HIROSHIMA MATH. J.
25 (1995), 171–205

A study of nonparametric estimation of error distribution in linear model based on L_1-norm

Zhuyu LI

(Received September 20, 1993)

1. Introduction

The field of nonparametric estimation has broadened its appeal in the last two decades with an array of new tools for statistical analysis. These new tools have offered sophisticated alternatives to the traditional parametric models in exploring large amounts of univariate or multivariate data without making any special distributional assumption. One of these tools is an estimation method of nonparametric density, which has become a prominent statistical research topic. Given identically distributed random variables $x_1, \cdots, x_n$ drawn from a population with density f, the aim is to construct an estimator of f without making any parametric assumption on the form of f. The pioneering papers might be due to Rosenblatt (1956) and Parzen (1962). Since the publication of these early papers, there has been a large amount of research on density estimation. In particular, theoretical and applied research on nonparametric density estimation has given noticeable influence on the related subjects, such as nonparametric regression, nonparametric discrimination, and so on, for the detail, see Alan (1991), Chao and Chai (1992), etc.

We consider a linear model

$$y_i = \boldsymbol{x}_i'\beta + e_i, \qquad i = 1, 2, \cdots, \tag{1.1}$$

where $\boldsymbol{x}_i'$s are $p(\geq 1)$-dimension known vectors and $\beta(\in R^p)$ is an unknown regression coefficient vector. The errors e_i's are assumed to be i.i.d. r.v.'s with a common unknown density function $f(x)$, and

$$E(e_1) = 0, \quad 0 < Var(e_1) = E(e_1^2) < \infty. \tag{1.2}$$

It is frequently assumed that e_1 has a normal distribution $N(0, \sigma^2)$ in usual regression analysis. Then, an estimator of β based on $(x_1, y_1), \cdots, (x_n, y_n)$ is obtained by the Least Squares method. The estimator, which is called the LSE of β, is defined as an unique solution $\hat{\beta}$ of the following minimization problem:

$$\sum_{i=1}^{n} (y_i - x_i'\hat{\beta})^2 = \min_{\beta \in R^p} \sum_{i=1}^{n} (y_i - x_i'\beta)^2. \tag{1.3}$$

It is known that the LSE $\hat{\beta}$ has some nice properties under the normality assumption. For a long time the LSE occupied a prominent position in the field of application. However, the normal assumption is unreasonable in some practical problems, and in such cases the behavior of the LSE is not as good enough as we can expect. For the case when the error distribution is unknown, we need to propose a good estimator for the error distribution and to consider a goodness-of-fit test for the error distribution.

Another motivation of the study of the error distribution is obtained by considering the Least Absolute Deviations (LAD) analysis (or be called minimum L_1-norm estimation) in linear models. Recently, the search work for robust procedures in statistical data analysis has generated considerable interest in developing statistical methods based on the LAD estimators, which use the L_1-norm rather than the L_2-norm. The LAD estimator $\hat{\beta}$ of β in the model (1.1) is defined as a Borel measurable solution of the minimization problem:

$$\sum_{i=1}^{n} |y_i - x_i'\tilde{\beta}| = \min_{\beta \in R^p} \sum_{i=1}^{n} |y_i - x_i'\beta| \tag{1.4}$$

under the condition

$$med(e_1) = 0 \tag{1.5}$$

instead of the LSE $\hat{\beta}$ of (1.3) under the conditions (1.2).

Historically, LAD estimation of unknown regression parameter vector β in model (1.1) dates back to Boscovich (1757) and Laplace (1793), but for a long time it has not attracted much attention. One reason is due to a computational difficulty, and the other is the lack of adequate theoretical studies on the LAD method. It is only recently that the LAD method has come into a prominence. This is due to phenomenal development in computational methods to solve complex optimization problems and robust techniques to deal with non-normal distributions and outliers. The computational problem, see Gentle, Narula and Sposito (1987), was successfully solved by linking the optimization problem with that of linear programming for which there are satisfactory algorithms. However, since no explicit form of $\hat{\beta}$ is available, a workable small sample theory has been little done. On the other hand, the asymptotic sampling theory of LAD estimators is now well developed both in univariate and multivariate linear models. For example, the basic problems of consistencies and asymptotic distributions of LAD estimators were studied by Bassett and Koenker (1978), Amemiya (1982), Bloomfield and

Steiger (1983), Ivanov (1984), McKean and Schrader (1987), Chen and Wu (1988), Rao (1988) and Wu (1988), Bai, Chen, Mia and Rao (1990), Bai, Rao and Yin (1990), Chen, Bai, Zhao and Wu (1990, 1992), Chen, Zhao and Wu (1993), etc. The LAD estimation method has been also used in a censored regression model that is widely applicable in econometrics, that is, in the model

$$y_i = \max\{0, \boldsymbol{x}_i'\beta + u_i\}, \qquad i = 1, 2, \cdots, n, \tag{1.6}$$

where the dependent variable y_i and the regression vector $\boldsymbol{x}_i$ are observed for each i, while the parameter vector β and error terms u_i are not observed. For such a model, the ordinary LSE for β is not appropriate, and an efficient censored least absolute deviation (LAD) estimator $\tilde{\beta}$ has been proposed by Powell (1984, 1986).

The asymptotic consistency and normality of $\tilde{\beta}$ have been established under some suitable regularity conditions. However, the problem of estimating $f(x)$ by using $\tilde{\beta}$ has been left. It may be noted that this problem is important because there are less conditions on the unobservable error terms e_i, $i = 1, 2, \cdots$, and hence we need to do regression diagnostics such as whether the model is appropriate or not. The asymptotic methods for testing linear hypotheses on β in model (1.1) based on LAD regression estimator have been recently discussed b a number of authors, but its statistics has been proposed by estimating an unknown quantity $f(0)$. Rao (1988) has noted that there is no satisfactory method of estimating $f(0)$. The same note has been done by Powell (1984), he notes: "The most difficult problem this poses-one which is generic to estimation methods based on least absolute deviations-is the estimation of the density function $f(\cdot)$ of the underlying error terms $\{e_i\}$". Furthermore, from these papers, we know that the asymptotic variance of $\tilde{\beta}$ depends on an unknown quantity $f(0)$, so, even the asymptotic theory is now well developed, but there is no satisfactory method of dealing with $f(0)$, or $\tau = [2f(0)]^{-1}$ (see McKean and Schrader (1987)). Therefore, the problem of estimating $f(x)$ is a very important work in min L_1-norm analysis.

In this paper, we propose a nonparametric method for estimating an unknown error distribution function $f(x)$ based on the LAD estimator in the general linear model (1.1) with condition (1.5), and prove that the nonparametric estimators have not only weak consistency, but also strong consistency. The asymptotic normality of the nonparametric estimator is also considered. The most difficulty in our study is: the estimator we propose here is based on residuals which are not independent, and also do not follow any fixed rules like those in dependent variables, say mixing variables. We use its symmetry to overcome this difficulty and the main technique here we use are similar to those in Chai, Li and Tian (1991) and Chai and Li (1993), in which we proposed a nonparametric procedure to estimate the unknown error distribution in model (1.1) by using the LSE. As our convenience in comparing these with the ones based on the LAD, we list some main results of those two papers in section 3, which will be used frequently in section 4. Some basic results in nonparametric kernel density estimators based on i.i.d. random variables will be listed in section 2. Our main theoretical results will be given in section 4, and some topics related to our density estimator will be given in section 5. Further, some computational exmples are given in section 6 to study a compares of two kinds of nonparametric density estimators of the unknown error distribution. It may be noted that asymptotic behaviors of the error distribution in the linear model (1.1) will be made clear by using the nonparametric kernel density estimators developed in the papers by Chai, Li and Tian (1991), Chai and Li (1993) and Zhang (1990) as well as the present paper.

2. Preliminaries

It is needless to say that nonparametric density estimators are recommended only if they possess desirable properties. In general, the research on them has settled on developing large sample properties. Some basic large sample properties of univariate kernel density estimators based on i.i.d. random variables have been given by many authors. The most basic works are based on the results of Parzen (1962). Here, we list some of these results, which will be used frequently in this paper.

Let $x_1, x_2, \cdots, x_n$ be independent and identically distributed random variables with a common density function f. Then the Rosenblatt-Parzen kernel estimator of f is of the form

$$f_n(x) = (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{x_i - x}{h_n}\right), \qquad x \in R^1, \tag{2.1}$$

where $K(\cdot)$ is a Borel measurable function on R^1 and $\{h_n\}$ is a sequence of positive numbers. The h_n is called window width. Some theoretical results are obtained for continuity points of f. Basic results on asymptotic unbiasedness and normality of $f_n(x)$ was established by Parzen (1962).

THEOREM 2.1. (*Asymptotic Unbiased and Asymptotic Normality of $f_n(x)$*). *Suppose that $K(\cdot)$ is a Borel measurable function satisfying the following conditions*:

(*i*) $\operatorname{Sup}_{-\infty<u<\infty} |K(u)| < \infty$;
(*ii*) $\int_{-\infty}^{\infty} |K(u)|\, du < \infty$;
(*iii*) $\lim_{u\to\infty} uK(u) = 0$;

(*iv*) $\int_{-\infty}^{\infty} K(u)du = 1$;

(*v*) $\lim_{n\to\infty} h_n = 0$.

Let x be a continuity point of x. Then the estimator $f_n(x)$ defined by (2.1) *is asymptotically unbiased, that is,*

$$\lim_{n\to\infty} E(f_n(x)) = f(x). \tag{2.2}$$

Further, $f_n(x)$ is also asymptotically normal, in the sense that, for every real number c,

$$\lim_{n\to\infty} P\left(\frac{f_n(x) - E(f_n(x))}{var\,(f_n(x))} \le c\right) = \Phi(c),$$

where Φ is the distribution function of the standard normal distribution given by

$$\Phi(c) = (2\pi)^{-1} \int_{-\infty}^{c} \exp\left\{-\frac{1}{2}y^2\right\} dy.$$

An important large sample property is to show asymptotic consistency of $f_n(x)$. The following basic and simplest results of consistency of $f_n(x)$ was also studied by Parzen (1962) with some more strong conditions on window width and unknown density function $f(x)$.

THEOREM 2.2. (*Consistent in Quadratic Mean of $f_n(x)$*). *Suppose that $K(\cdot)$ and h_n satisfy the conditions of Theorem* 2.1, *and* $\lim_{n\to\infty} nh_n = \infty$. *Then*

$$\lim_{n\to\infty} E[f_n(x) - f(x)]^2 = 0. \tag{2.3}$$

From Theorem 2.2 it follows that $f_n(x)$ has a weak consistency, that is, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|f_n(x) - f(x)| \ge \varepsilon) = 0. \tag{2.4}$$

THEOREM 2.3. (*Pointwise Strong Consistency of $f_n(x)$*). *Suppose that $K(\cdot)$ and h_n satisfy the conditions of Theorem* 2.1, *and* $\lim_{n\to\infty} nh_n/\log n = \infty$. *Then at every point of continuity of $f(\cdot)$,*

$$\lim_{n\to\infty} f_n(x) = f(x), \qquad a.s. \tag{2.5}$$

THEOREM 2.4. (*Strong Uniform Consistency of $f_n(x)$*). *Suppose that f is uniformly continuous on R^1, and $K(\cdot)$ and h_n satisfy the conditions of Theorem* 2.3, *then*

$$\lim_{n\to\infty}\{\sup_x |f_n(x)-f(x)|\}=0, \qquad a.s. \tag{2.6}$$

Some extension of the above results to the multivariate case has been studied. Based on the i.i.d. random vectors $\boldsymbol{x}_i$, an extended multivariate kernel density estimator is defined by

$$f_n(\boldsymbol{x})=(nh_n^d)^{-1}\sum_{j=1}^n K\left(\frac{\boldsymbol{x}-\boldsymbol{x}_j}{h_n}\right), \qquad \boldsymbol{x}\in R^d.$$

For the basic results of this estimator, see Chapter 3.1 of Prakasa Rao (1983).

3. Nonparametric density estimate of $f(x)$ based on LSE $\hat{\beta}$

The nonparametric density estimator of error distribution was extended by Chai, Li and Tian (1991), Chai and Li (1993) to the case of model (1.1) with the condition (1.2). The estimator is based on the LSE $\hat{\beta}$ of β. In section 4 we introduce an alternative estimator, which is based on the LAD estimator $\tilde{\beta}$ of β. One of the main purposes in this in this paper is to study some basic asymptotic properties of the latter estimator. In the study of these properties, we also use some results on the density estimator based on the LSE $\hat{\beta}$, and so some basic results are stated in this section. This will be also useful in comparison with two estimators based on the LSE $\hat{\beta}$ and the LAD estimator $\tilde{\beta}$.

Consider the linear model (1.1), where the errors e_i's are assumed to be i.i.d. random variables with a common unknown density function $f(x)$, and satisfy the conditions (1.2). Let $\hat{\beta}$ be the LSE of β based on $(x_1, y_1),\cdots,(x_n, y_n)$, which is given by (1.3). Let $\hat{e}_i$ be the residuals based on $\hat{\beta}$, that is,

$$\hat{e}_i=y_i-\boldsymbol{x}_i'\hat{\beta}. \tag{3.1}$$

Then the Rosenblatt estimator of $f(x)$ is defined as

$$\hat{f}_n(x)=(nh_n)^{-1}\sum_{i=1}^n K\left(\frac{\hat{e}_i-x}{h_n}\right). \tag{3.2}$$

Some theoretical properties of this estimator have been obtained only when the kernel function $K(\cdot)$ has a special form as

$$K(x)=\frac{1}{2}I_{[-1,1)}=\begin{cases}\frac{1}{2}, & \text{if } -1\le x<1,\\ 0, & \text{otherwise.}\end{cases} \tag{3.3}$$

In this case, since here $K(\cdot)$ is a discrete function, we can rewrite (3.2) as

$$\hat{f}_n(x)=(2h_n)^{-1}[\hat{F}_n(x+h_n)-\hat{F}_n(x-h_n)], \qquad x\in R^1,$$

or

$$\hat{f}_n(x) = (2nh_n)^{-1} \sum_{i=1}^{n} I_{(x-h_n \le \hat{e}_i \le x+h_n)}, \qquad x \in R^1, \tag{3.4}$$

where $\hat{F}_n$ denotes the empirical distribution function of $\hat{e}_1, \hat{e}_2, \cdots, \hat{e}_n$ and $I_{(\cdot)}$ is an indicator function.

We list some basic conditions restricted to all $\boldsymbol{x}_i$'s and $f(x)$:

(i) There exists a constant $M > 0$ such that $\|\boldsymbol{x}_i\| \le M$ for all $i \ge 1$, where $\|\cdot\|$ is an Euclidean norm in R^p;

(ii) $\lim_{n\to\infty} \frac{1}{n} S_n = \Sigma > 0$, where $S_n = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'$;

(iii) $f(x) > 0$ at x;

(iv) $f(x)$ satisfies the local Lipschitz's condition at x, that is, there exist constants $d = d(x) > 0$ and $\delta = \delta(x) > 0$ dependent only on x such that

$$|f(t) - f(x)| \le d\,|t - x|, \quad \text{whenever} \quad t \in (x - \delta,\, x + \delta).$$

THEOREM 3.1. (*Weak Consistency of* $\hat{f}_n(x)$). *Let* $e_1, \cdots, e_n$, *be i.i.d. r.v.'s with density* f *in model* (1.1), *and suppose that* (1.2) *holds and* $\boldsymbol{x}_i'$*s satisfy the above conditions* (*i*) *and* (*ii*). *If*

$$h_n \longrightarrow 0 \quad \textit{and} \quad \sqrt{n}\, h_n \longrightarrow \infty, \quad \textit{as} \quad n \longrightarrow \infty,$$

then

$$\hat{f}_n(x) \xrightarrow{P} f(x), \quad \textit{as} \quad n \longrightarrow \infty, \tag{3.5}$$

where $\hat{f}_n(x)$ *is defined by* (3.2), $x \in C(f)$ *and* $C(f)$ *is the set of continuity points of* $f(\cdot)$.

THEOREM 3.2. (*Point-wise Strong Consistency of* $\hat{f}_n(x)$). *Under the conditions in Theorem* 3.1, *if*

$$h_n \longrightarrow 0 \quad \textit{and} \quad \frac{\sqrt{n}\, h_n}{\log n} \longrightarrow \infty, \quad \textit{as} \quad n \longrightarrow \infty,$$

then

$$\hat{f}_n(x) \longrightarrow f(x) \quad \textit{a.s.} \quad \textit{as } n \longrightarrow \infty, \tag{3.6}$$

where $x \in C(f)$.

THEOREM 3.3. (*Uniform Strong Consistency of* $\hat{f}_n(x)$). *Under the conditions of Theorem* 3.2, *if* f *uniform continuous on* R^1, *and*

$$h_n \longrightarrow 0 \quad and \quad \frac{\sqrt{n}h_n}{\log n} \longrightarrow \infty, \quad as \quad n \longrightarrow \infty,$$

then

$$\sup_x |\hat{f}_n(x) - f(x)| \longrightarrow 0 \quad a.s. \quad as \ n \longrightarrow \infty. \tag{3.7}$$

The proofs of Theorems 3.1, 3.2 and 3.3 can be seen in Chai, Li and Tian (1991).

THEOREM 3.4. (L_1-*Norm Consistency of* $\hat{f}_n(x)$). *If*

$$\lim_{n\to\infty} h_n = 0 \quad and \quad \inf_n \frac{\sqrt{n}h_n}{\log n} > 0,$$

then under the basic conditions (*i*) *and* (*ii*) *it holds that for any* $\varepsilon > 0$ *and any* $f \in \mathscr{F}$,

$$P(|\hat{f}_n(x) - f(x)|_{L_1} > \varepsilon) \le c \exp\{-cn\varepsilon^2\}, \tag{3.8}$$

where $\mathscr{F}$ *is the set of the density functions on* R^1, *and* c *is a positive constant such that it does not dependent on* n *and also the parameters in model* (1.1).

THEOREM 3.5. (*Asymptotic Normality of* $\hat{f}_n(x)$). *If*

$$\lim_{n\to\infty} nh_n^3 = 0 \quad and \quad \lim_{n\to\infty} \frac{n^{5/12}h_n}{\log n} = \infty,$$

then under the basic conditions (*i*)-(*iv*) *it holds that*

$$\sqrt{\frac{2nh_n}{f(x)}}\, |\hat{f}_n(x) - f(x)| \xrightarrow{L} N(0, 1), \quad as \quad n \longrightarrow \infty. \tag{3.9}$$

THEOREM 3.6. (*Law of Iterated Logarith on* $\hat{f}_n(x)$). *Under the conditions of Theorem* 3.5, *if*

$$\lim_{n\to\infty} \frac{nh_n^3}{\log\log n} = 0 \quad and \quad \lim_{n\to\infty} \frac{\sqrt{n}h_n \log\log n}{(\log n)^2} = \infty,$$

then

$$\lim_{n\to\infty} \sup_n \pm \sqrt{\frac{nh_n}{f(x)\log\log n}}\, [\hat{f}_n(x) - f(x)] = 1 \qquad a.s. \tag{3.10}$$

The proofs of Theorems 3.4, 3.5 and 3.6 can be found in Chai and Li (1993). It may be noted that the above results have been proved only for the special kernel estimator (3.3). For the general kernel estimator of error distribution based on LSE, Zhang [51] has obtained certain asymptotic results which are analogous to the results of Theorems 3.1, 3.2 and 3.3.

4. Asymptotic theory of estimation of $f(x)$ based on L_1-norm in linear model

In this section, we propose a nonparametric method to estimate an unknown density function $f(x)$ by using the general kernel function based on alternative residuals in a linear model. In model (1.1), unlike the usual linear model, we do not assume the existence of the moments of e_i, but we assume only that its median is equal to zero, that is,

$$med\,(e_1) = 0. \tag{4.1}$$

Obviously, this model will be useful in the many practical problems, since we make a weak assumption on the model. Following (3.1), consider the alternative residuals defined by

$$\tilde{e}_i = y_i - \boldsymbol{x}_i' \tilde{\beta}, \qquad i = 1, 2, \cdots. \tag{4.2}$$

Let $K(\cdot)$ be a Borel measurable function on R^1 satisfying the conditions in Theorem 2.1. Then the kernel estimator $\tilde{f}_n(x)$ of $f(x)$ based on $\tilde{e}_i$ is defined as

$$\tilde{f}_n(x) = (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{\tilde{e}_i - x}{h_n}\right), \qquad x \in R^1, \tag{4.3}$$

where $h_n > 0$ is a window width constant. Obviously, if $K(\cdot)$ is continuous, $\tilde{f}_n(x)$ is a continuous estimation function of $f(x)$ with respect to x. We will stusy the asymptotic properties of $\tilde{f}_n(x)$ in (4.3) in this section.

First of all, we give some lemmas used in the following.

4.1. Some lemmas

The following inequality for the sum of i.i.d. variables is frequently used in studying asymtotic properties of nonparametric density estimators.

LEMMA 4.1. (*Bernstein Inequality*). *Let $Z_1, Z_2, \cdots, Z_n$ be independent r.v.'s satisfying $|Z_i| \leq c$, a.s. and $E(Z_i) = 0$, $i = 1, 2, \cdots, n$. Then for any $t > 0$ and $n \geq 1$,*

$$P(|\sum_{i=1}^{n} Z_i| \geq nt) \leq 2 \exp\{-n^2 t^2/[2 \sum_{i=1}^{n} var\,(Z_i) + (2/3)c]\}, \tag{4.4}$$

where c is an appropriate constant.

LEMMA 4.2. (*Devroye and Wagner* (1980)). *Let* μ_n *and* μ *be the empirical and theoretical distribution functions, respectively. Further, let* $I_{(A)}$ *be an interval with length* $A > 0$. *Then for any* $\varepsilon > 0$ *and* $A > 0$ *such that* $0 < \mu(I_{(A)}) \le b \le 1/4$,

$$P(\sup\{|\mu_n(I_{(A)}) - \mu(I_{(A)})| \ge \varepsilon\}) \le 16n^2 \exp\left\{-\frac{n\varepsilon^2}{64b + 4\varepsilon}\right\} + 8n \exp\{-nb/10\}. \tag{4.5}$$

The following rsults are basic and important for the asymptotic theoretical properties of LAD estimator $\tilde{\beta}$ defined by (1.4).

LEMMA 4.3. (*Chen, Bai, Zhao and Wu* (1992)). *In model* (1.1) *with* (4.1), *suppose that there exist constants* $l_1 > 0$ *and* $l_2 > 0$ *such that*

$$P(0 < e_i < h) \ge l_2 h \ \&\ P(-h < e_i < 0) \ge l_2 h, \quad \textit{whenever} \quad h \in (0, l_1), \tag{4.6}$$

and under the assumption that S_n^{-1} *exists, let*

$$d_n = \max_{1 \le i \le n} x_i' S_n^{-1} x_i. \tag{4.7}$$

Then the following assertions hold:

(1) $d_n = o(1) \Rightarrow \tilde{\beta} \xrightarrow{P} \beta$;

(2) $d_n = o(1/\log n) \Rightarrow \tilde{\beta} \to \beta$ *a.s.*

(3) $d_n = O(1/n) \Rightarrow \tilde{\beta}$ *tends to* β *exponentially in the sense that* $P(\|\tilde{\beta} - \beta\| \ge \varepsilon) = O(e^{-cn})$ *for any* $\varepsilon > 0$, *with a constant* c *independent of* n *but possibly dependent on* ε.

LEMMA 4.4. *Under the same conditions as in Lemma* 4.3, *we can get that*

$$d_n = O(1/n) \Longrightarrow \|\tilde{\beta}_n - \beta\| = O(n^{-1/2} \log n) \qquad a.s. \tag{4.8}$$

PROOF. The proof can be obtained by slightly modifying the proof of Lemma 4.3 as in the following.

Let $x_i^{*'} = S_n^{-1/2} x_i$, $i = 1, 2, \cdots, n$, and $\beta_0^* = S_n^{1/2}\beta$. Then model (1.1) can be expressed as

$$y_i = x_i^{*'} \beta_0^* + e_i, \qquad 1 \le i \le n. \tag{4.9}$$

Denote the LAD estimator of β_0^* by $\tilde{\beta}^*$ in model (4.9). Then we have

$$\tilde{\beta} = S_n^{-1/2}\tilde{\beta}^*, \ \sum_{i=1}^n x_i^* x_i^{*'} = I_p, \qquad d_n = \max_{1 \le i \le n} \|x_i^*\|^2, \tag{4.10}$$

$$\|S_n^{-1}\| = \|P_n' \Lambda_n^{-1} P_n\| \le \underline{\lambda}^{-1}(S_n), \ \|S_n^{-1/2}\tilde{\beta}^*\| \le \|\tilde{\beta}^*\| \cdot \|S_n^{-1/2}\|,$$

where $\|A\| = \max_{i,j}\{|a_{i,j}|\}$ for a matrix A, $\|x_i\| = (\sum_{i=1}^n x_i' x_i)^{1/2}$ *and* $\underline{\lambda}(A)$ denotes the smallest eigenvalue of matrix A.

From the proof of Lemma 4.3, for any $\varepsilon > 0$,

$$\begin{aligned} P(\|\tilde{\beta} - \beta\| \geq \varepsilon n^{-1/2} \log n) &\leq P(\|\tilde{\beta}^*\| \geq \varepsilon n^{-1/2} \log n \sqrt{\underline{\lambda}(S_n)}) \\ &\leq P(\|\tilde{\beta}^*\| \geq \varepsilon n^{-1/2} \log n \sqrt{c_0/d_n}) \\ &\leq P(\|\tilde{\beta}^*\| \geq V_n), \end{aligned}$$

where $V_n = O(\log n)$. Further, from the formula (2.1) in Chen, Bai, Zhao and Wu (1992), we obtain that the last inequality is upper bounded by $\exp\{-cV_n^2\}$, that is,

$$P(\|\tilde{\beta} - \beta\| \geq \varepsilon n^{-1/2} \log n) \leq \exp\{-c \log^2 n\}.$$

Thus, we get (4.8).

LEMMA 4.5. (*Chen, Bai, Zhao and Wu* (1990)). *In model* (1.1), *suppose that e_i's are i.i.d. with a common density function $f(x)$, and* (4.1) *is satisfied. Further, we assume the following two conditions*:

(*i*) *There exists a constant $\Delta > 0$, such that $f(u) = F'(u)$ for $|u| \leq \Delta$, $f(0) > 0$ and $f(x)$ is continuous at $x = 0$;*

(*ii*) *$d_n \to 0$ as $n \to \infty$, where d_n is defined by* (4.7).

Then

$$2f(0)S_n^{1/2}(\tilde{\beta} - \beta) \xrightarrow{L} N(0, I_p), \qquad as\ n \longrightarrow \infty.$$

COROLLARY 4.1. *If $(1/n)S_n \to \Sigma$, as $n \to \infty$, then, under the conditions of Lemma* 4.5 *it holds that*

$$\sqrt{n}(\tilde{\beta} - \beta) \xrightarrow{L} N(0, [2f(0)]^{-2}\Sigma^{-1}), \qquad as\ n \longrightarrow \infty. \tag{4.11}$$

4.2. Main results

In the following, we will study the large sample properties of $\tilde{f}_n(x)$. First of all, we start our study from the asymptotic consistency.

LEMMA 4.6. *In model* (1.1), *suppose that e_i's satisfy the conditions of Lemma* 4.5, *and also the following conditions are satisfied*:

(1) $\lim_{n\to\infty} \dfrac{1}{n} S_n = \Sigma > 0$,

(2) $h_n \longrightarrow 0$, $\sqrt{n}h_n \longrightarrow \infty$, *as* $n \longrightarrow \infty$.

Then for a 'true' parameter vector β_0 *and for any positive constant* $a > 0$, *it holds that for* $x \in C(f)$,

$$(nh_n)^{-1} \sum_{i=1}^{n} I_{(x+ah_n \le e_i \le x+ah_n+M\,\|\tilde{\beta}-\beta_0\|)} \xrightarrow{P} 0, \qquad \text{as } n \longrightarrow \infty. \tag{4.12}$$

PROOF. For a fixed $x \in C(f)$, let

$$J_n^*(x) = (nh_n)^{-1} \sum_{i=1}^{n} I_{(x+ah_n \le e_i \le x+ah_n+M\,\|\tilde{\beta}-\beta_0\|)}\,. \tag{4.13}$$

From Corollary of Lemma 4.5, for any $\varepsilon > 0$ there exists an $\eta > 0$ such that

$$P(\sqrt{n}\,\|\tilde{\beta} - \beta_0\| > \eta) < \varepsilon,$$

for large n. Therefore, for any $\varepsilon_0 > 0$

$$\begin{aligned} P(J_n^*(x) > \varepsilon_0) &\le P(\sqrt{n}\,\|\tilde{\beta} - \beta_0\| > \eta) + P(J_n^*(x) > \varepsilon_0,\ \sqrt{n}\,\|\tilde{\beta} - \beta_0\| \le \eta) \\ &\le \varepsilon + P(J_n^*(x) > \varepsilon_0,\ \sqrt{n}\,\|\tilde{\beta} - \beta_0\| \le \eta) \\ &\le \varepsilon + P(\sum_{i=1}^{n} I_{(x+ah_n \le e_i \le x+ah_n+M\eta/\sqrt{n})} > nh_n\varepsilon_0). \end{aligned} \tag{4.14}$$

Noting that $x \in C(f)$, and using Chebyshev inequality, it can be seen that the second term in (4.14) is upper bounded by

$$(h_n\varepsilon_0)^{-1} P(x + ah_n \le e_1 \le x + ah_n + M\eta/\sqrt{n}) = \frac{M\eta}{\sqrt{n}\,h_n\varepsilon_0} f(x)(1 + 0(1)).$$

Therefore, from the condition (2) it follows that

$$\lim_{n\to\infty} \sup_{x} P(J_n^*(x) > \varepsilon_0) \le \varepsilon,$$

and hence, $J_n^*(x) \xrightarrow{P} 0$ as $n \to \infty$. So, we get (4.12). □

THEOREM 4.1. (*Weak Consistency of* $\tilde{f}_n(x)$). *Suppose that the conditions of Lemma* 4.5, 4.6 *and the following conditions are satisfied*:

(1) *There exists a constant* $M > 0$, *such that* $\|\boldsymbol{x}_i\| \le M$,

(2) $K(\cdot)$ *is a bounded Riemann integrable p.d. function on* R^1,

(3) *There exists a constant* $\rho > 0$ *such that* $K(u) = 0$, *for* $|u| > \rho$.

Let x *be a continuous point of* f. *Then*

$$\tilde{f}_n(x) \xrightarrow{P} f(x), \qquad \text{as } n \longrightarrow \infty. \tag{4.15}$$

PROOF. For a fixed $x \in C(f)$, let

$$f_n(x) = (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{e_i - x}{h_n}\right).$$

Then we have

$$|\tilde{f}_n(x) - f(x)| \le |\tilde{f}_n(x) - f_n(x)| + |f_n(x) - f(x)|,$$

and from Theorem 1.2, we know

$$P(|f_n(x) - f(x)| \ge \varepsilon) \longrightarrow 0, \qquad \text{as } n \longrightarrow \infty.$$

So we need only to prove that

$$P(|\tilde{f}_n(x) - f_n(x)| \ge \varepsilon) \longrightarrow 0, \qquad \text{as } n \longrightarrow \infty. \tag{4.16}$$

Suppose that K is a non-negative, bounded Riemann integrable function on R^1. Then, by Lemma 3 of Devroye and Wagner (1980), for each ε, δ, $\rho > 0$, we can find a function $K^*(u)$ such that

$$K^*(u) = \sum_{i=1}^{N} a_i I_{A_i}(u), \qquad u \in R^1, \tag{4.17}$$

where I_{A_i} is the indicator function, and

(i) $a_1, \cdots, a_N$ are non-negative real numbers,
(ii) $A_1, \cdots, A_N$ are disjoint intervals contained in $[-\rho, \rho]$,
(iii) $K^*(u) \le \sup_u K(u) = k^*$, $u \in R^1$,
(iv) $|K^*(u) - K(u)| < \varepsilon$ on $[-\rho, \rho]$ except on a set D,
(v) $D \subseteq B = \bigcup_1^M B_i$ where $B_1, \cdots, B_M$ are intervals from $[-\rho, \rho]$, whose union has Lebesgue measure less than δ, where μ is the Lebesgue measure on R^1.

We can write

$$\begin{aligned} J &= |\tilde{f}_n(x) - f_n(x)| \\ &= h_n^{-1} \left| \int K\left(\frac{u-x}{h_n}\right) d\hat{\mu}_n(u) - \int K\left(\frac{u-x}{h_n}\right) d\mu_n(u) \right| \\ &\le \sum_{i=1}^{3} J_i. \end{aligned}$$

Here

$$J_1 = h_n^{-1} \int \left| K^*\left(\frac{u-x}{h_n}\right) - K\left(\frac{u-x}{h_n}\right) \right| d\mu_n(u),$$

$$J_2 = h_n^{-1} \int \left| K^*\left(\frac{u-x}{h_n}\right) - K^*\left(\frac{u-x}{h_n}\right) \right| d\hat{\mu}_n(u),$$

$$J_3 = h_n^{-1} \left| \int K^*\left(\frac{u-x}{h_n}\right) d\hat{\mu}_n(u) - \int K^*\left(\frac{u-x}{h_n}\right) d\mu_n(u) \right|,$$

where $\hat{\mu}_n$ and μ_n denote the empirical distributions of $\tilde{e}_i$ and e_i respectively, and μ is the distribution function of e_i.

Use the same notations as in Chai (1984):

$$D_2 = S_{(x,\rho h_n)} \cap (x + h_n D), \quad D_3 = S_{(x,\rho h_n)} \cap (x + h_n D)^c, \tag{4.18}$$

$$S_{(x,\rho h_n)} = \{u \colon |u - x| \le \rho h_n\}, \quad x + h_n D = \{x + h_n a \colon a \in D\}. \tag{4.19}$$

Noting that $x \in C(f)$, we can that there exists a constant $\delta_0 > 0$ and a constant $M_0 > 0$ such that when $|h_n t| < \delta_0$, $t \in D, \Rightarrow |f(x + th_n)| \le M_0$, where M_0 may depend on x. Then we obtain

$$\begin{aligned}
J_1 &\le 2k^* h_n^{-1} \int_{D_2} d\mu_n(u) + \varepsilon h_n^{-1} \int_{D_3} d\mu_n(u) \\
&\le 2k^* h_n^{-1}(\mu(D_2) + |\mu_n(D_2) - \mu(D_2)|) + \varepsilon h_n^{-1}(\mu(D_3) + |\mu_n(D_3) - \mu(D_3)|) \\
&\le 2k^* M_0 \delta + 2\varepsilon M_0 \rho \\
&\quad + 2k^* h_n^{-1} |\mu_n(D_2) - \mu(D_2)| + \varepsilon h_n^{-1} |\mu_n(D_3) - \mu(D_3)|.
\end{aligned} \tag{4.20}$$

Similarly

$$\begin{aligned}
J_2 &\le 2k^* h_n^{-1} \int_{D_2} d\hat{\mu}_n(u) + \varepsilon h_n^{-1} \int_{D_3} d\hat{\mu}_n(u) \\
&\le 2k^* h_n^{-1}(\mu(D_2) + |\hat{\mu}_n(D_2) - \mu(D_2)|) + \varepsilon h_n^{-1}(\mu(D_3) + |\hat{\mu}_n(D_3) - \mu(D_3)|) \\
&\le 2k^* M_0 \delta + 2\varepsilon M_0 \rho \\
&\quad + 2k^* h_n^{-1} |\mu_n(D_2) - \mu(D_2)| + \varepsilon h_n^{-1} |\mu_n(D_3) - \mu(D_3)| \\
&\quad + 2k^* h_n^{-1} |\hat{\mu}_n(D_2) - \mu_n(D_2)| + \varepsilon h_n^{-1} |\hat{\mu}_n(D_3) - \mu_n(D_3)|.
\end{aligned} \tag{4.21}$$

Now we consider

$$\begin{aligned}
J_3 &= (nh_n)^{-1} \left| \sum_{j=1}^{n} \sum_{i=1}^{N} a_i [I_{(\tilde{e}_j \in (x + A_i h_n))} - I_{(e_j \in (x + A_i h_n))}] \right| \\
&\le \sum_{i}^{N} a_i \left[(nh_n)^{-1} \sum_{j=1}^{n} |I_{(\tilde{e}_j \in (x + A_i h_n))} - I_{(e_j \in (x + A_i h_n))}|\right].
\end{aligned} \tag{4.22}$$

Noting that $A_i = (a_i, b_i) \subset [-\rho, \rho]$ are disjoint intervals, $a_i < b_i$, $a_i \ne b_i$, we have

$$
\begin{aligned}
&(nh_n)^{-1}\textstyle\sum_{j=1}^n |I_{(\tilde{e}_j\in(x+A_ih_n))} - I_{(e_j\in(x+A_ih_n))}| \\
&\quad = (nh_n)^{-1}\sum_{j=1}^n (I_{(\tilde{e}_j\in(x+A_ih_n))\cap(e_j\in(x+A_ih_n)^c)} + I_{(\tilde{e}_j\in(x+A_ih_n)^c)\cap(e_j\in(x+A_ih_n))} \\
&\quad \le (nh_n)^{-1}\sum_{j=1}^n I_{(x+a_ih_n-M\|\tilde{\beta}-\beta_0\|<e_j\le x+a_ih_n)} \\
&\qquad + (nh_n)^{-1}\sum_{j=1}^n I_{(x+a_ih_n<e_j\le x+a_ih_n+M\|\tilde{\beta}-\beta_0\|)} \\
&\qquad + (nh_n)^{-1}\sum_{j=1}^n I_{(x+b_ih_n-M\|\tilde{\beta}-\beta_0\|\le e_j<x+b_ih_n)} \\
&\qquad + (nh_n)^{-1}\sum_{j=1}^n I_{(x+b_ih_n\le e_j<x+b_ih_n+M\|\tilde{\beta}-\beta_0\|)}.
\end{aligned}
\tag{4.23}
$$

From Lemma 4.6 it follows that the four terms in (4.23) converge to zero in probability as $n\to\infty$. This implies $J_3 \xrightarrow{P} 0$ as $n\to\infty$.

Now we consider the terms appeared in the last expressions of (4.20) and (4.21). Note that

$$
\begin{aligned}
h_n^{-1}|\hat{\mu}_n(D_2)-\mu_n(D_2)| &= (nh_n)^{-1}|\sum_{j=1}^n (I_{(\tilde{e}_j\in D_2)} - I_{(e_j\in D_2)})| \\
&\le (nh_n)^{-1}\sum_{j=1}^n |I_{(\tilde{e}_j\in D_2)} - I_{(e_j\in D_2)}|,
\end{aligned}
\tag{4.24}
$$

and

$$
|I_{(\tilde{e}_j\in D_2)} - I_{(e_j\in D_2)}| = I_{(\tilde{e}_j\in D_2)\cap(e_j\in D_2^c)} + I_{(\tilde{e}_j\in D_2^c)\cap(e_j\in D_2)}. \tag{4.25}
$$

Since D_2 is an union of the disjoint intervals $\subset[-\rho,\rho]$, we can make the same statement as in (4.23) to get

$$
h_n^{-1}|\hat{\mu}_n(D_2)-\mu_n(D_2)| \xrightarrow{P} 0. \tag{4.26}
$$

Similarly, it holds that

$$
h_n^{-1}|\hat{\mu}_n(D_3)-\mu_n(D_3)| \xrightarrow{P} 0. \tag{4.27}
$$

From Lemma 4.1, we know that for any $\varepsilon>0$,

$$
\begin{aligned}
P\{(nh_n)^{-1}|\sum_{i=1}^n [I_{(e_i\in D_2)} - P(e_i\in D_2)]| \ge \varepsilon\} &= P(|\sum_{i=1}^n (I_{(e_i\in D_2)} - P(e_i\in D_2))| \ge \varepsilon nh_n) \\
&\le 2\exp\{-Cnh_n\varepsilon\},
\end{aligned}
\tag{4.28}
$$

where C is an appropriate positive constant. Because the last expression in (4.28) is upper bounded by $C/n^{-\alpha}$ $(\alpha > 1)$ for large n, it holds that

$$\sum_{n=1}^{\infty} \exp\{-Cnh_n\varepsilon\} < \infty, \tag{4.29}$$

and hence, by Borel-Cantelli lemma, we get

$$h_n^{-1}|\mu_n(D_2) - \mu(D_2)| \xrightarrow{a.s.} 0, \qquad \text{as } n \longrightarrow \infty. \tag{4.30}$$

Similarly we can prove

$$h_n^{-1}|\mu_n(D_3) - \mu(D_3)| \xrightarrow{a.s.} 0, \qquad \text{as } n \longrightarrow \infty. \tag{4.31}$$

So, $J_1 \xrightarrow{P} 0$, $J_2 \xrightarrow{P} 0$, as $n \to \infty$ and this completes the proof of the theorem. □

LEMMA 4.7. *Assume that the conditions of Lemmas* 4.3, 4.5 *and* 4.6 *are satisfied. If*

$$h_n \longrightarrow 0 \text{ and } \sqrt{n}h_n/\log n \longrightarrow \infty, \qquad \text{as } n \longrightarrow \infty \tag{4.32}$$

and $d_n = O(1/n)$, *then for any* $a > 0$ *and* $x \in C(f)$,

$$(nh_n)^{-1} \sum_{i=1}^{n} I_{(x+ah_n < e_i \le x+ah_n+M\|\tilde{\beta}-\beta_0\|)} \xrightarrow{a.s.} 0, \qquad \text{as } n \longrightarrow \infty. \tag{4.33}$$

PROOF. From lemma 4.4, we know that if $d_n = O(1/n)$, then

$$\|\tilde{\beta} - \beta_0\| = O(n^{-1/2}\log n) \qquad a.s.$$

Fix an $x \in C(f)$, and let

$$f_n^*(x) = h_n^{-1}P(x + ah_n \le e_i \le x + ah_n + Cn^{-1/2}\log n). \tag{4.34}$$

Then, since x is a continuous point of f,

$$f_n^*(x) \le Cf(x)(n^{1/2}h_n/\log n)^{-1},$$

when n is large. From condition (4.32), we obtain

$$f_n^*(x) \longrightarrow 0, \qquad \text{as } n \longrightarrow \infty. \tag{4.35}$$

Let

$$A_n = \{\|\tilde{\beta} - \beta_0\| \ge Cn^{-1/2}\log n\}.$$

Then from Lemma 4.4 it holds that

$$P(A_n) \le \exp\{-C\log^2 n\} < \varepsilon_0,$$

when n is large. Using (4.13), we have

$$P(J_n^*(x) > \varepsilon) = P(J_n^* > \varepsilon,\ \|\tilde{\beta} - \beta_0\| < Cn^{-1/2}\log n) + P(A_n)$$
$$\le \varepsilon_0 + P((nh_n)^{-1}\sum_{i=1}^{n} I_{(x+ah_n \le e_i \le x+ah_n+Cn^{-\frac{1}{2}}\log n)}).$$

Let

$$V_i = I_{(x+ah_n \le e_i \le x+ah_n+Cn^{-\frac{1}{2}}\log n)} - P(x+ah_n \le e_i \le x+ah_n+Cn^{-1/2}\log n),$$

$i = 1, 2, \cdots, n$. Then V_i are independent, $|V_i| \le 1$, $E(V_i) = 0$, $1 \le i \le n$ and $var(V_i) \le Cf(x)n^{-1/2}\log n$, for large n. Therefore from Lemma 4.1 and condition (4.32), it follows that for any $\varepsilon > 0$

$$P(|J_n^*(x) - f_n^*(x)| > \varepsilon) \le P(|\sum_{i=1}^{n} V_i| \ge nh_n\varepsilon)$$
$$\le 2\exp\{-Cnh_n\varepsilon^2/(C(\log n/\sqrt{n}h_n)f(x) + \varepsilon)\}$$
$$\le 2\exp\{-Cnh_n\varepsilon\}.$$

Applying Borel-Cantelli lemma to this result, we get

$$J_n^*(x) - f_n^*(x) \longrightarrow 0 \quad a.s. \quad \text{as } n \longrightarrow \infty. \tag{4.36}$$

From (4.35) and (4.36), we obtain $J_n^*(x) \to 0$ a.s. as $n \to \infty$, which implies (4.33).

From Lemma 4.7 and (4.23), we can get, as $n \to \infty$

$$(nh_n)^{-1}\sum_{j=1}^{n}|I_{(\tilde{e}_j \in (x+A_i h_n))} - I_{(e_j \in (x+A_i h_n))}|$$
$$\le (nh_n)^{-1}\sum_{j=1}^{n} I_{(x+a_i h_n - M\|\tilde{\beta}-\beta_0\| < e_j \le x+a_i h_n)}$$
$$+ (nh_n)^{-1}\sum_{j=1}^{n} I_{(x+a_i h_n < e_j \le x+a_i h_n + M\|\tilde{\beta}-\beta_0\|)}$$
$$+ (nh_n)^{-1}\sum_{j=1}^{n} I_{(x+b_i h_n - M\|\tilde{\beta}-\beta_0\| \le e_j < x+b_i h_n)}$$
$$+ (nh_n)^{-1}\sum_{j=1}^{n} I_{(x+b_i h_n \le e_j < x+b_i h_n + M\|\tilde{\beta}-\beta_0\|)}$$
$$\xrightarrow{a.s.} 0.$$

This means $J_3 \xrightarrow{a.s.} 0$ as $n \to \infty$. By the arguments similar to the above method, we can get

$$(nh_n)^{-1}|\hat{\mu}_n(D_2) - \mu_n(D_2)| \xrightarrow{a.s.} 0$$

and

$$(nh_n)^{-1}|\hat{\mu}_n(D_3) - \mu_n(D_3)| \xrightarrow{a.s.} 0$$

as $n \to \infty$.

Finally, combining these results to (4.30) and (4.31), we get

$$\tilde{f}_n(x) - f_n(x) \longrightarrow 0 \quad a.s. \quad \text{as } n \longrightarrow \infty,$$

which is summarized as follows:

THEOREM 4.2 (*Pointwise Strong Consistency of* $\tilde{f}_n(x)$). *Suppose that the conditions of Theorem* 4.1 *are satisfied. Then, if* $h_n \to 0$ *and* $\sqrt{n}h_n/\log n \to \infty$ *as* $n \to \infty$, *it holds that for* $x \in C(f)$

$$\tilde{f}_n(x) \longrightarrow f(x), \quad a.s. \quad as\ n \to \infty. \tag{4.37}$$

LEMMA 4.8. *In the addition to the conditions of Lemma* 4.7, *suppose that* f *is uniformly continuous on* R^1. *Then*

$$\sup_x (nh_n)^{-1} \sum_{i=1}^n I_{(x+ah_n \le e_i \le x+ah_n+M\|\tilde{\beta}-\beta_0\|)} \xrightarrow{a.s.} 0, \qquad as\ n \longrightarrow \infty. \tag{4.38}$$

PROOF. From the proof of Lemma 4.7, it is vident that $\sup_x f_n^*(x) \to 0$, as $n \to \infty$. So, to prove (4.38), we need only to prove

$$\sup_x |J_n^*(x) - f_n^*(x)| \longrightarrow 0, \quad a.s. \quad \text{as } n \longrightarrow \infty. \tag{4.39}$$

Note $\sup_x f(x) = f_0 < \infty$, which follows from the uniform continuity of f. So, when n is large, we obtain $0 < c_n = f_0 M b_n < 1/4$, and

$$\sup \mu([x + ah_n,\ x + ah_n + Mb_n]) \le c_n, \tag{4.40}$$

where $b_n = Cn^{-1/2} \log n$.

By Lemma 4.2, with $A_n^c = \{\|\tilde{\beta} - \beta_0\| < Cn^{-1/2} \log n\}$, it holds that for any $\varepsilon > 0$ and large n

$$\begin{aligned} &P(\sup_x |J_n^*(x) - f_n^*(x)| \ge \varepsilon) \\ &\quad \le P(\sup_x (nh_n)^{-1} | \sum_{i=1}^n I_{(x+ah_n \le e_i \le x+ah_n+Mb_n)} \\ &\qquad - \sum^n P(x + ah_n \le x + ah_n + Mb_n)| \ge \varepsilon) \end{aligned}$$

$$\le P(\sup_x |\mu_n(x + ah_n,\ x + ah_n + Mb_n) - \mu(x + ah_n,\ x + ah_n + Mb_n)| \ge \varepsilon h_n)$$
$$\le 16n^2 \exp\{-nh_n^2\varepsilon^2/(64Cb_n + 4h_n\varepsilon)\} + 8n\exp\{-nCb_n/10\}$$
$$\le 16n^2 \exp\{-nh_n\varepsilon^2/(64C\log n/\sqrt{n}h_n + \varepsilon)\} + 8n\exp\{-C\sqrt{n}\log n\}.$$

Therefore, using Borel-Cantelli lemma, we have

$$\sup_x |J_n^*(x) - f_n^*(x)| \longrightarrow 0 \quad a.s. \quad \text{as } n \longrightarrow \infty,$$

which proves (4.39), and hence (4.38). □

THEOREM 4.3. *Under the conditions of Theorem* 4.2, *if* $f(x)$ *is uniformly continuous on* R^1, *then*

$$\sup_x |\tilde{f}_n(x) - f(x)| \longrightarrow 0, \qquad a.s. \tag{4.41}$$

PROOF. From Theorem 2.4, we know

$$\sup_x |f_n(x) - f(x)| \longrightarrow 0 \quad a.s. \quad \text{as } n \longrightarrow \infty.$$

So it is sufficient to prove that

$$\sup_x |\tilde{f}_n(x) - f_n(x)| \longrightarrow 0 \quad a.s., \quad \text{as } n \longrightarrow \infty, \tag{4.42}$$

that is,

$$\sup_x J = \sup_x |\tilde{f}_n(x) - f_n(x)| \le \sup \sum_{i=1}^{3} J_i \longrightarrow 0.$$

We use the same reductions as in the proof of Theorem 4.1. Let S be the set of all intervals in R^1, and $s = [b, d] \in S$. Then from (4.17), (4.22) and (4.23) we have

$$\sup_x J_3 = \sup_x h_n^{-1} \left| \int K^*((u - x)/h_n) d\hat{\mu}_n(u) - \int K^*((u - x)/h_n) d\mu_n(u) \right|$$
$$\le Nk^* \sup_{s\in S} h_n^{-1} |\hat{\mu}_n(s) - \mu_n(s)|$$
$$= Nk^* \sup_{s\in S} \{(nh_n)^{-1} | \sum_{j=1}^{n} [I_{(\tilde{e}_j \in s)} - I_{(e_j \in s)}|\}$$
$$\le Nk^* \sup_{s\in S} \{(nh_n)^{-1} \sum_{i=1}^{n} |I_{(\tilde{e}_j \in s)} - I_{(e_j \in s)}|\}$$

$$\leq Nk^* \sup_b \left\{(nh_n)^{-1} \sum_{j=1}^{n} I_{(b-M\|\tilde{\beta}-\beta_0\| \leq e_j < b)}\right\}$$

$$+ Nk^* \sup_b \left\{(nh_n)^{-1} \sum_{j=1}^{n} I_{(b \leq e_j < b+M\|\tilde{\beta}-\beta_0\|)}\right\}$$

$$+ Nk^* \sup_d \left\{(nh_n)^{-1} \sum_{j=1}^{n} I_{(d-M\|\tilde{\beta}-\beta_0\| \leq e_j < d)}\right\}$$

$$+ Nk^* \sup_d \left\{(nh_n)^{-1} \sum_{j=1}^{n} I_{(d < e_j \leq d+M\|\tilde{\beta}-\beta_0\|)}\right\}. \tag{4.43}$$

So, by Lemma 4.8, we can get

$$\sup_x J_3 \longrightarrow 0 \quad a.s. \quad \text{as } n \longrightarrow \infty. \tag{4.44}$$

Since $f(x)$ is uniformly continuous on R^1, there exists a constant $M_1 > 0$ such that $|f(x)| < M_1$. Let

$$M^* = \max\{M_0, M_1\}.$$

Then, from (4.20) and (4.21) in the proof of Theorem 4.1 we obtain

$$\begin{aligned}\sup_x J_1 + \sup_x J_2 \leq\ & 4k^* M^* \delta + 4\varepsilon M^* \rho \\ & + (4k^*/h_n) \sup_x |\mu_n(D_2) - \mu(D_2)| \\ & + (2\varepsilon/h_n) \sup_x |\mu_n(D_3) - \mu(D_3)| \\ & + (2k^*/h_n) \sup_x |\hat{\mu}_n(D_2) - \mu_n(D_2)| \\ & + (\varepsilon/h_n) \sup_x |\hat{\mu}_n(D_3) - \mu_n(D_3)|. \end{aligned} \tag{4.45}$$

By the arguments similar to (4.44), we get

$$h_n^{-1} \sup_x |\hat{\mu}_n(D_2) - \mu_n(D_2)| \xrightarrow{a.s.} 0 \tag{4.46}$$

and

$$h_n^{-1} \sup_x |\hat{\mu}_n(D_3) - \mu_n(D_3)| \xrightarrow{a.s.} 0 \tag{4.47}$$

as $n \to \infty$.

Using Dvoretsky-Kiefer-Wolfowitz inequality (see Dvoretsky, Kiefer and Wolfowitz (1956)) and (28) and (29) in Chai (1984), we have

$$h_n^{-1} \sup_x |\mu_n(D_2) - \mu(D_2)| \xrightarrow{a.s.} 0 \tag{4.48}$$

and

$$h_n^{-1} \sup_x |\mu_n(D_3) - \mu(D_3)| \xrightarrow{a.s.} 0 \tag{4.49}$$

as $n \to \infty$. Therefore, (4.42) follows from the arbitrariness of ε, δ and (4.44) $\sim$ (4.49). $\square$

THEOREM 4.4 *Suppose that the conditions of Theorem* 4.2 *are satisfied. Then it holds that* (*i*) *if* $f(x)$ *satisfies the local Lipschitz condition at* x, *and* $x \in C(f)$, *then*

$$\tilde{f}_n(x) - f(x) = O(n^{-1/4} \log n); \tag{4.50}$$

and (*ii*) *if* $f(x)$ *is uniformly continuous on* R^1,

$$\sup_x |\tilde{f}_n(x) - f(x)| = O(n^{-1/4} \log n) \qquad a.s. \tag{4.51}$$

For the definition of the local Lipschitz condition, see the basic conditions (iv) in section 3. The results (4.50) and (4.51) can be proved by the method similar to arguments as in the proofs of Theorem 4.2 and 4.3 and letting $h_n = n^{-1/4}$.

In the following we will give the asymptotic normality of $\tilde{f}_n(x)$ when $K(\cdot)$ has the special form given by

$$K(x) = (1/2)I_{(-1,1]} = \begin{cases} \dfrac{1}{2}, & \text{if } -1 < x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

This is a special case of (4.3). In this case, we can simply rewrite $\tilde{f}_n(x)$ as

$$\tilde{f}_n(x) = (2h_n)^{-1} \sum_{i=1}^{n} I_{(x-h_n < \tilde{e}_i \leq x+h_n)}, \qquad x \in R^1. \tag{4.52}$$

First we state two important lemmas.

LEMMA 4.9 (Chai and Li (1993)). *If* $\lim_{n\to\infty} nh_n^3 = 0$, *then*

$$\sqrt{2nh_n/f(x)}\,(f_n(x)-f(x)) \xrightarrow{L} N(0,\,1),$$

where $f_n(x)$ *is defined by* (2.1).

LEMMA 4.10 (Chai and Li (1993)). *If*

$$\lim_{n\to\infty} h_n = 0 \quad and \quad \lim_{n\to\infty} n^{5/12} h_n/\log n = \infty,$$

then as $n \to \infty$,

$$(1/\sqrt{nh_n}) \sup_{0<m_i\le c_n} |\sum_{i=1}^{n} (I_{(x\pm h_n,\, x\pm h_n+m_i)}(e_i) - \mu(x\pm h_n,\, x\pm h_n+m_i))| = o_p(1),$$

for $i = 1, 2, \cdots, n$, *where* $c_n = Cn^{-1/2}$.

THEOREM 4.5 (*Asymptotic Normality of* $\tilde{f}_n(x)$). *Suppose that the basic conditions on* $\boldsymbol{x}_i$*'s and* $f(x)$ *given by* (*i*) $\sim$ (*iv*) *in section* 3 *are satisfied and* $f(0) > 0$. *If*

$$\lim_{n\to\infty} nh_n^3 = 0 \quad and \quad \lim_{n\to\infty} n^{5/12} h_n/\log n = \infty,$$

then

$$\sqrt{2nh_n/f(x)}\,|\tilde{f}_n(x)-f(x)| \xrightarrow{L} N(0,\,1). \tag{4.53}$$

PROOF. To prove (4.53), it is sufficient to show that

$$\sqrt{2nh_n/f(x)}\,(f_n(x)-f(x)) \longrightarrow N(0,\,1) \tag{4.54}$$

and

$$\sqrt{nh_n}\,(\tilde{f}_n(x)-f_n(x)) = o_p(1). \tag{4.55}$$

The first result follows from Lemma 4.9. The second result can be proved by the same way as in the proof of Theorem 3.5, and so we give only a brief sketch here. Using the Corollary of Lemma 4.5, for any $\varepsilon > 0$ there exists a $\eta > 0$ such that

$$P(\max_{1\le i\le n} \sqrt{n}\,|\boldsymbol{x}_i'(\tilde{\beta}-\beta)| > \eta) < \varepsilon \tag{4.56}$$

for large n. Since (4.56) is the same result as the (3.7) in Chai and Li (1993), so we could use Lemma 4.10 to get (4.55) directly.

3.4. Remarks

First we examine certain differences between the conditions restricted to asymptotic theory of two types of density estimators $\hat{f}_n(x)$ and $\tilde{f}_n(x)$ which are based on LS and LAD methods, respectively. From Theorem 4.1 $\sim$ 4.5 we can note that the basic conditions on error terms e_i in model (1.1) for obtaining asymptotic properties of $\tilde{f}_n(x)$ are those from Lemma 4.3 and 4.5, that is

$$med\,(e_1) = 0$$

and there exist constants $l_1 > 0$ and $l_2 > 0$ such that

$$P(0 < e_i < h) \geq l_2 h \ \ \& \ \ P(-h < e_i < 0) \geq l_2 h,$$

for all $h \in (0, l_1)$, $i = 1, 2, \cdots, n$. The latter condition requires that there should be enough concentration of probability of e_i in the vicinity of median zero instead of certain assumptions on moments of e_i. This can be gone back to LAD estimator $\tilde{\beta}$ of β in (1.1) with condition (4.1). From Chen, Zhao and Wu (1993), we can see that the key assumption in the asymptotic theory of $\tilde{\beta}$ is a local behavior of the distribution of e_i in the vicinity of 0. On the other hand, for obtaining asymptotic properties of $\hat{f}_n(x)$, it required that e_i's satisfy certain moment conditions as in (1.2). Further, from section 3, we know that asymptotic properties of $\hat{f}_n(x)$ only require $f(x) > 0$ at x. Therefore, we can ignore the behavior of e_i about the point zero when we hold those conditions for e_i's moments.

Conditions on the design points are the same for both two kinds of nonparametric estimators in model (1.1), that is, the condition

$$\lim \frac{1}{n} S_n > 0, \quad S_n = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'$$

plays an important role in the asymptotic theory of the estimator $\tilde{f}_n(x)$ as well as the estimator $\hat{f}_n(x)$.

Next we see a relationship between the density estimators of the observations y_i's and the error e_i's. For this, consider the simplest linear model given by

$$y_i = \beta + e_i, \quad i = 1, \cdots, n, \tag{4.57}$$

where the errors e_i's are i.i.d. r.v. with a common unknown density function f and β is unknown. This is a special case of model (1.1). Then y_i's are i.i.d. r.v.'s with the density

$$g(y) = f(y - \beta). \tag{4.58}$$

Based on the sample of size n, we have an estimator of g,

$$\hat{g}_n(y) = (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{y_i - x}{h_n}\right), \qquad x \in R^1.$$

On the other hand, we have proposed two types of estimators $\hat{f}_n$ and $\tilde{f}_n$ for f given by (3.2) and (4.3), respectively. Here we note that there exists a natural relationship between these estimators, that is, given by

$$\begin{aligned}\hat{g}_n(y) &= \hat{f}_n(y - \hat{\beta}) \\ &= \tilde{f}_n(y - \tilde{\beta}).\end{aligned}$$

Various asymptotic properties of $\hat{f}_n(y - \hat{\beta})$ or $\tilde{f}_n(y - \tilde{\beta})$ follow from the results in section 2.

Lastly, from the present paper, Chai, Li and Tian (1991) and Chai and Li (1993) it is possible to use asymptotic consistent nonparametric estimators of an unknown error density function f in the linear model (1.1) based on both two kinds estimators of β. The error density estimation can be used to check on appropriateness of the model, i.e. it will give a direct impact on diagnostics of the model. Therefore, estimating $f(x)$ in a nonlinear regression model is a more important problem. However, this problem seems to have been little treated. Ahmad (1992) noted this problem and gave his consideration about estimation of $f(x)$, but the main formula (A.3) he used in his proofs seems to be not correct. Naturally, we hope we can extend our results to a nonlinear regression model. This will be our future work.

5. Some related topics

In this paper, we have proposed an estimator $\tilde{f}_n(x)$ for f in model (1.1) with condition (4.1). This will be useful in the diagnoses of model (1.1). Further we note that we need to estimate $f(0)$ since the asymptotic covariance matrix of $\tilde{\beta}$ involves a unknown $f(0)$. In this section we see the other cases in which the estimator $\tilde{f}_n(x)$ based on LAD estimator are useful.

5.1. Estimation of Error Distribution in the LAD Test

In model (1.1) with condition (4.1) we consider a general linear hypothesis

$$H_0\colon A\beta = 0, \tag{5.1}$$

where A is a $q \times p$ matrix of rank q. Let

$$B = \min_{A\beta=\alpha} \sum_{i=1}^{n} |y_i - \boldsymbol{x}_i'\beta| - \min_{\beta\in R^p} \sum_{i=1}^{n} |y_i - \boldsymbol{x}_i'\beta|.$$

Then by an analogy with the least squares theory, it has been proposed to use a test statistic

$$F = \frac{B}{q\tilde{\tau}/2}, \tag{5.2}$$

where $\tilde{\tau}$ is an estimator of $\tau = [2f(o)]^{-1}$. So we need to get at least one consistent estimator of $f(0)$. The fact that τ is related to the test statistic will be understood as follows. Bai, Rao and Yin (1987) gave an important result on asymptotic distribution of B, that is, under H_0

$$B = f(0)(\hat{\beta} - \tilde{\beta})' S_n(\hat{\beta} - \tilde{\beta}) + o_p(1). \tag{5.3}$$

Therefore using Lemma 4.5, it holds that under H_0

$$4f(0)B \xrightarrow{L} \chi^2(m), \tag{5.4}$$

which requires to find an estimator of $f(0)$. Rao (1988) mentioned that several suggestions have been made to estimate $f(0)$ consistently. But he notes that there is no satisfactory method of estimating $f(0)$. In view of difficulties in estimating $f(0)$ in usual way, Rao gave the way to estimate $f(0)$ by rewrite model (1.1) when the sample size is sufficiently large. Here, we suggest to use $\tilde{f}_n(0)$ as an estimator of $f(0)$.

From (4.3), we can get the estimator of $f(0)$ as

$$\tilde{f}_n(0) = (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{\tilde{e}_i}{h_n}\right). \tag{5.5}$$

The following is obtained as a special result of Theorem 4.1, 4.2 and 4.4.

THEOREM 5.1. *Suppose that* $h_n \to 0$ *as* $n \to \infty$. *Then under conditions of Theorem* 4.1 *it holds that*

(1) *if* $\sqrt{n}h_n \to \infty$, $\tilde{f}_n(0) \xrightarrow{P} f(0)$;

(2) *if* $\sqrt{n}h_n/\log n \to \infty$, $\tilde{f}_n(0) \to f(0)$ *a.s.*

(3) *if* $\sqrt{n}h_n/\log n \to \infty$ *and* $h_n = n^{-1/4}$,

$$\tilde{f}_n(0) - f(0) = O(n^{-1/4} \log n).$$

5.2. Nonparametric estimator of the scale parameter for the rank analysis of model (1.1)

There has been a considerable development in robust methods of linear models by using rank statistics. The details can be seen in Chapter 4 of [17], Jackel (1972), Koul, Sievers and McKean (1987). Under model (1.1), the rank estimator of β is defined as the values of β which minimizes Jackel's

(1972) dispersion function given by

$$D_\varphi(\beta) = \sum_{i=1}^{n} \alpha(R(y_i - x_i'\beta))(y_i - x_i'\beta), \tag{5.6}$$

where $R(z_i)$ denotes the rank of z_i among $z_1,\cdots,z_n$ and the rank scores are generated as $\alpha(i) = \varphi\left(\frac{i}{n+1}\right)$ for a non decreasing function φ defined on (0, 1). In this rank analysis it is important to estimate the scale parameter

$$\gamma = \gamma(f) = \int f d\varphi(F) \tag{5.7}$$

in the model (1.1). Note $\gamma = \int \varphi'(F(x))f^2(x)dx$ when φ is differentiable. The estimator of γ is used to stangardize test statistics. The quantity γ also appears in comparisons of rank procedures, and so it is of interest to estimate γ. Some estimators have been proposed in Koul, Sievers and McKean (1987). In this section we propose an alternative estimator based on $\hat{f}_n(x)$ (Or similarly, based on $\tilde{f}_n(x)$). Using (24) in Koul, Sievers and McKean (1987), an estimator of the scale parameter γ can be made as

$$\hat{\gamma}_n = \int \hat{f}_n(x) d\varphi(F_n(x)) = \sum (\varphi(i/n) - \varphi((i-1)/n))\hat{f}_n. \tag{5.8}$$

Here, F_n is the empirical d. f. of $\hat{e}_i$, $i = 1, 2,\cdots,n$, and $\hat{f}_n(x)$ is the same as (3.2). We make the following assumptions:

(1) There exist $S_n^{-1} = (\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i')^{-1}$ for all $n \geq p$ and

$$d_n = \max_{1\leq i\leq n} \boldsymbol{x}_i' S_n^{-1} \boldsymbol{x}_i \longrightarrow 0.$$

(2) $F(x)$ has a uniformly continuous, bounded density function f(x), $f(x) > 0$ a.e., $f(x) \to 0$ as $x \to \pm\infty$.

(3) There exists a $M > 0$, such that $\|\boldsymbol{x}_i\| \leq M$ for any i.

It may be noted that conditions (1) and (2) are the same as (A1) and (A2) in Koul, Sievers and McKean (1987), but condition (3) is much weaker than (A3). In addition to a condition on window width constant h_n, that is,

$$h_n \longrightarrow 0 \quad \text{and} \quad \sqrt{n}h_n/\log n \longrightarrow \infty, \quad \text{as} \quad n \longrightarrow \infty,$$

then we have the following theorem.

THEOREM 5.2. *Under the assumptions mentioned above it holds that*

$$\sup_{\varphi\in C} |\hat{\gamma}_n - \gamma(f)| \longrightarrow 0 \qquad a.s. \tag{5.9}$$

where C is a class of score functions satisfying $C = \{\varphi: [0, 1] \to [0, 1],\ \varphi$ non decreasing, right continuous and bounded, $\varphi(0) = 0,\ \varphi(1) = 1\}$.

PROOF. From (5.7) and (5.8)

$$\begin{aligned}
\sup_{\varphi\in C} |\hat{\gamma}_n - \gamma| &= \left| \int \hat{f}_n(x) d\varphi(F_n(x)) - \int f(x) d\varphi(F(x)) \right| \\
&= \left| \int_0^1 \hat{f}_n(F_n^{-1}(t)) d\varphi(t) - \int_0^1 f(F^{-1}(t)) d\varphi(t) \right| \\
&\leq \sup_t |\hat{f}_n(F_n^{-1}(t)) - f(F^{-1}(t))| \\
&\leq \sup_t |\hat{f}_n(F_n^{-1}(t)) - f(F_n^{-1}(t))| + \sup_t |f(F_n^{-1}(t)) - f(F^{-1}(t))| \\
&\leq \sup_y |\hat{f}_n(y) - f(y)| + \sup_t |f(F_n^{-1}(t)) - f(F^{-1}(t))|,
\end{aligned}$$

where for each $t \in [0, 1]$,

$$F_n^{-1}(t) = \inf\{x: F_n(x) \geq t\}.$$

By using a transformation

$$\int_{-\infty}^{+\infty} F(x) d\varphi(F(x)) = \int_0^1 t d\varphi(t),$$

we can see that the first term in the last expression convergences to 0 a.s. From the condition (2) in the above, we get (5.9). □

If $f(x)$ satisfies the local Lipschitz's condition at x we get the following theorem.

THEOREM 5.3. *Suppose that the conditions of Theorem 5.2 are satisfied. Then*

$$\sup_{\varphi\in C} |\hat{\gamma}_n - \gamma(f)| = O(n^{-1/4} \log n), \qquad a.s. \tag{5.10}$$

PROOF. From the proof of Theorem 5.2 and the Corollary of Theorem 4 in Chai, Li and Tian (1991), it follows that

$$\sup_x |\hat{f}_n(x) - f(x)| = O(n^{-1/4} \log n) \qquad \text{a.s.}$$

and hence we get (5.10) immediately by using the same reductions as in the proofs of Theorem 5.1. □

6. Numerical examples

In this section we give some examples to illustrate behaviors of two types of estimators for the density of error distribution in a linear regression model. One is

$$\tilde{f}_n(x) = (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{\tilde{e}_i - x}{h_n}\right), \tag{6.1}$$

which is based on the least absolute deviation estimator (LAD) of model (1.1) with the median condition (1.5); and the other is

$$\hat{f}_n(x) = (nh_n)^{-1} \sum_{i=1}^{n} K\left(\frac{\hat{e}_i - x}{h_n}\right), \tag{6.2}$$

which is based on the least square estimator (LSE) of model (1.1) with the moment conditions (1.2). We employ the regression model that was used to describe the relationship between the heights of the old generation and the younger generation in Chai, Li and Tian (1991). Let x be the height of the older generation and y be that of the younger generation. Suppose that the n independent observations (x_i, y_i), $i = 1, \cdots, n$ on (x, y) follow a linear regression model:

$$y_i = a + bx_i + e_i, \qquad i = 1, \cdots, n, \tag{6.3}$$

where $a = 85.6742$, $b = 5.16$, and e_i are unobservable errors with an unknown density function $f(x)$. Here we consider two cases when (i) $f(x)$ is the normal density function on $N(0, 1/4)$, and (ii) $f(x)$ is the density function of the student t-distribution $T(2)$ with 2 degree of freedom. Note that both densities are symmetric about zero, but no moments exist for the second case. In order to see behaviors of two estimators, we made new observations on y for each x_i by using the relation (6.3) and producing samples from the two error distributions. So, we have the following two new regression models:

$$y_{ij} = a_j + b_j x_i + e_{ij}, \qquad i = 1, \cdots, n, \tag{6.4}$$

where $j = 1, 2$, e_{i1}'s are random samples from $N(0, 1/4)$, and e_{i2}'s are random samples from $T(2)$. From now on, in (6.4) it is assumed that a_j and b_j are unknown. Further, we take the sample size n as 30 and 150. First we obtain the LS and LAD estimators of a_j and b_j, which are denoted by $\hat{a}_j$, $\hat{b}_j$ and $\tilde{a}_j$, $\tilde{b}_j$, respectively. These values are given in Tables 1. Here the LAD estimators were obtained by using the Fortran program in Armstrong and

Kung (1978). The r denotes the correlation between the observed values y_{ij} and the predicted values $\hat{y}_{ij}$ or $\tilde{y}_{ij}$, where $\hat{y}_{ij}$ and $\tilde{y}_{ij}$ are computed by using $(\hat{a}_j, \hat{b}_j)$ and $(\tilde{a}_j, \tilde{b}_j)$, respectively.

Table 1. LS and LAD estimators

f(x)	sample size	estimation method	a	b	r
T(2)	n = 30	LS	97.0267	4.4549	0.7886
		LAD	89.0014	4.9445	0.8133
T(2)	n = 150	LS	84.1474	5.2578	0.8765
		LAD	86.0427	5.1413	0.8717
N(0, 1/4)	n = 30	LS	89.756	4.8957	0.9803
		LAD	86.9946	5.0677	0.9809
N(0, 1/4)	n = 150	LS	87.3449	5.0481	0.9786
		LAD	86.5222	5.0992	0.9789

Now with two residuals

$$\hat{e}_{ij} = \hat{y}_{ij} - y_i \quad \text{and} \quad \tilde{e}_{ij} = \tilde{y}_{ij} - y_i, \tag{6.5}$$

we can easily compute (6.1) and (6.2) with a given kernel function $K(\cdot)$ and a window width h_n. Here, we employed the following four different kernel density functions in our study.

(i) uniform kernel function:

$$K_1(x) = \begin{cases} \frac{1}{2}, & \text{if } |x| \le 1, \\ 0, & \text{if } |x| > 1. \end{cases} \tag{6.6}$$

(ii) parabola type kernel function:

$$K_2(x) = \begin{cases} \frac{3}{4}(\sqrt{5})^{-3}((\sqrt{5})^2 - x^2), & \text{if } x^2 \le (\sqrt{5})^2, \\ 0, & \text{if } x^2 > (\sqrt{5})^2. \end{cases} \tag{6.7}$$

(iii) cosine type kernel function:

$$K_3(x) = \begin{cases} \frac{1}{2}\cos x, & \text{if } -\pi/2 \le x < \pi/2, \\ 0, & \text{otherwise}. \end{cases} \tag{6.8}$$

These three kernel functions, with bounded support, hold all requirements of kernel function $K(\cdot)$ for the asymptotic properties of $\tilde{f}_n(x)$ and $\hat{f}_n(x)$ we discussed in section 3 and 4. Since a popular choice of univariate kernels is the normal kernel function, we also include it in our numerical study, even it has an unbounded support.

(iv) normal type kernel function:

$$K_4(x) = (2\pi)^{-1/2} \exp\left\{-\frac{1}{2}x^2\right\}, \qquad -\infty < x < \infty. \tag{6.9}$$

Letting $K(x) = K_i(x)$, $i = 1, 2, 3, 4$ in (6.1) and (6.2), all different $\tilde{f}_n(x)$, $\hat{f}_n(x)$ can be easily calculated at 41 test points in the range of $x \in [-2, 2]$ for the sample sizes $n = 30$ and $n = 150$. We set window width h_n as $h_n = n^{-1/4}$ for $N(0, 1/4)$ and $h_n = n^{-1/5}$ for $T(2)$. These are given in Figures $1 \sim 16$. Here we use the following abbreviations and notations:

From these figures, we can see that if these errors come from a normal distribution, two nonparametric estimators show almost the same roles for fitting the error density function. Roughly, we can say that cosine kernel shows better smoothing than the other two kinds of bounded kernels. On the other hand, when errors come from $T(2)$ distribution and n is small, $\hat{f}_n(x)$ shows some left bias about the center of the true density function $T(2)$ for all four kinds kernels and $\tilde{f}_n(x)$ is much better. We can see that the probability of $\tilde{f}_n(x)$ concentrates about the center of true density function. An interesting point is appeared for the cases of using K_4, that is, Figure 4, 8, 12 and 16. They all show considerable good fitting and smoothing to the true density function even they do not meet the condition of bounded support in our theorems. This might suggest a need to extend our theoretical results to the case of kernels with certain unbounded support. We also can note that the choice of the window width h_n is also a topic left for further work.

Acknowledgments

The author would like to express her heartfelt gratitude to Professor Yasunori Fujikoshi, Hiroshima University, for his valuable comments and discussions; and also to Professor Genxiang Chai, Tongji University, for his constructive suggesting and valuable guidance all over this work. The author would like to show her sincere thanks to Professor Minoru Shiotani, for his continuous encouragement and his guidance to statistics. A special thanks goes to Dr. Jian Shi for his encouragement and support during the writing of this paper. This research was partially supported by National Natural Science Foundation of China Grant 18971061.

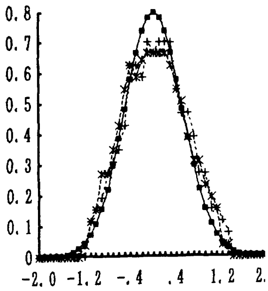


Figure 1. $K_1(x)$

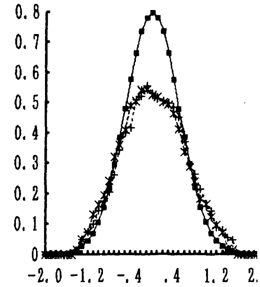
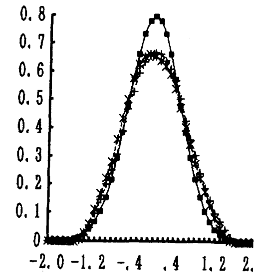


Figure 2. $K_2(x)$



Firure 3. $K_3(x)$

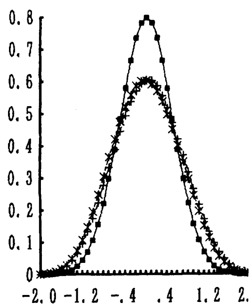


Figure 4. $K_4(x)$

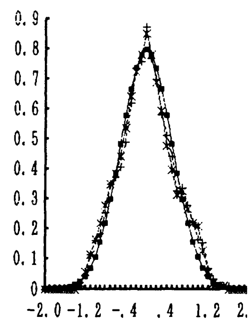


Figure 5. $K_1(x)$

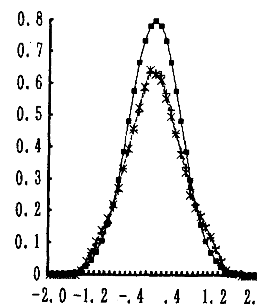
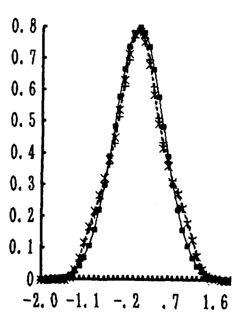


Figure 6. $K_2(x)$



Firure 7. $K_3(x)$

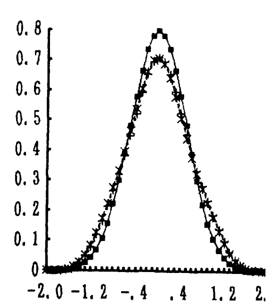


Figure 8. $K_4(x)$

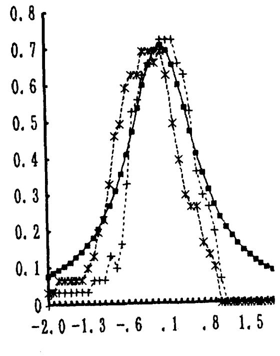


Figure 9. $K_1(x)$

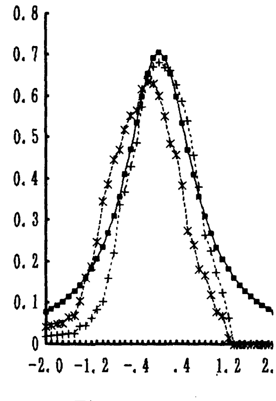
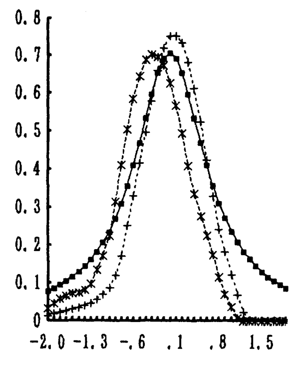


Figure 10. $K_2(x)$



Firure 11. $K_3(x)$

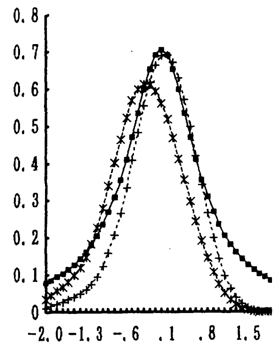


Figure 12. $K_4(x)$

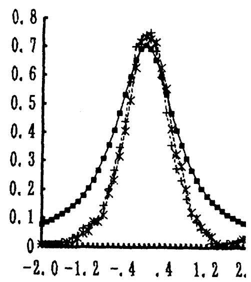


Figure 13. $K_1(x)$

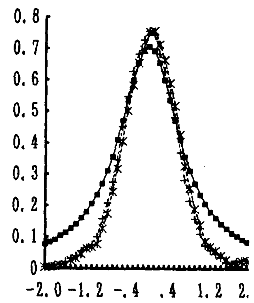
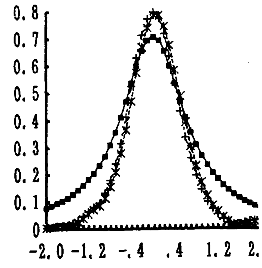


Figure 14. $K_2(x)$



Firure 15. $K_3(x)$

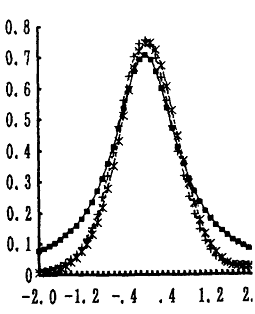


Figure 16. $K_4(x)$

Figures 1 ~ 4: $e_i \sim N(0, 1/4)$, $n = 30$.
Figures 5 ~ 8: $e_i \sim N(0, 1/4)$, $n = 150$.
Figures 9 ~ 12: $e_i \sim T(2)$, $n = 30$.
Figures 13 ~ 16: $e_i \sim T(2)$, $n = 150$.
—●—●—: $f(x)$
—*—*—: $\hat{f}_n(x)$
—+—+—: $\tilde{f}_n(x)$

References

[1] I. A. Ahmad, Residuals density estimation in nonparametric regression, Statistics & Probability Letters, **14** (1992), 133–139.

[2] J. I. Alan, Recent developments in nonparametric density estimation, J. Amer. Statist. Assoc., **86** (1991), 205–224.

[3] T. Amemiya, Two stage least absolute deviations estimators, Econometrica, **50** (1982), 689–711.

[4] R. D. Armstrong and M. T. Kung, Least absolute value estimates for a simple linear regression problem, Appl. Statist., **27** (1978), 363–366.

[5] A. Azzalini and A. Bowman, On the use of nonparametric regression for the checking linear relationships, J. R. Statist. Soc., **B 55** (1993), 549–557.

[6] G. J. Babu and C. R. Rao, Joint asymptotic distribution of marginal quantile functions in samples from a multivariate population, J. Multivariate Analysis, **27** (1988), 15–23.

[7] G. J. Babu, Strong representations for LAD estimators in linear models, Probab. Th. Rel. Fields, **83** (1989), 547–558.

[8] Z. D. Bai, X. R. Chen, B. Q. Mia and C. R. Rao, Asymptotic theory fo least distances estimate in multivariate models, Statistics, **21** (1990), 503–519.

[9] Z. D. Bai, C. R. Rao and Y. Q. Yin, Least absolute deviations analysis of variance, Sankhya, **A 52** (1990), 166–177.

[10] G. J. Bassett and R. Koenker, Asymptotic theory of least absolute error regression, J. Amer. Statist. Assoc., **73** (1978), 618–622.

[11] P. Bloomfiedld and W. L. Steiger, Least Absolute Deviations: Theory, application, and algorithms, Birkhauser, Boston, MA, 1983.

[12] G. X. Chai, Consistent estimation of random window-width kernel of density function, Scientia Sinica, Ser. A, English Ed., **27** (1984), 1153–1163.

[13] G. X. Chai, Z. Y. Li and H. Tian, Consistent nonparametric estimation of error distributions in linear model, ACTA Mathemeticae Applicatae Sinica, **7** (1991), 245–256.

[14] G. X. Chai and Z. Y. Li, Asymptotic theory for estimation of error distribution in linear model, Science in China, Ser. A, English Ed., **36** (1993), 408–419.

[15] X. R. Chen, Uniform convergence rates of kernel density estimation, J. Systems Science and Math. Sciences, **3** (1983), 263–272.

[16] X. R. Chen, G. J. Chen, Q. G. Wu and L. C. Zhao, Theory of Parameter Estimation in Linear Model, Science Press, Beijing, 1985 (in Chinese).

[17] X. R. Chen, Z. B. Fan, G. Y. Li and B. Tao, Nonparametric Statistics, Science Press, Shanghai, 1989 (in Chinese).

[18] X. R. Chen, Z. D. Bai, L. C. Zhao and Y. H. Wu, Asymptotic normality of minimum L_1-norm estimates in linear models, Science in China, Ser. A, English Ed., **33** (1990), 1311–1328.

[19] X. R. Chen, Z. D. Bai, L. C. Zhao and Y. H. Wu, Consistency of minimum L_1-norm estimates in linear model, The development of Statistics: recent contributions from China, Pitman Research Notes in Math. Series, **258** (1992), 249–260.

[20] X. R. Chen and Y. H. Wu, On a necessary condition for the consistency of the L_1 estimates in linear regression models, Commun. Statist.-Theory Meth., **22** (1993), 631–639.

[21] X. R. Chen and Y. H. Wu, On consistency of minimum L_1-norm estimates in linear models, In Probability and Statistics: Proceedings of the Special Program at Nankai Institute of Mathematics, 32–38, World Scientific, Singapore, 1992.

[22] X. R. Chen, L. C. Zhao and Y. H. Wu, On conditions of consistency of ML_1N estimates, Statistica Sinica, **3** (1993), 9–18.

[23] L. P. Devroye and T. J. Wagner, The strong uniform consistency of kernel density estimates, Multivariate Analysis **V**, Ed. by P. R. Kirshnaiah, North-Holland Publishing Company (1980), 59–77. 59–77.

[24] L. P. Devroye, The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates, Ann. Statist., **11** (1983), 896–904.

[25] A. Dvoretsky, J. Kiefer and J. Wolfowitz, Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, Ann. Math. Statist., **27** (1956), 642–669.

[26] J. E. Gentle, S. C. Narula and V. A. Sposito, Algorithms for the un constrained L_1 linear regression, In Statistical Data Analysis Based on the L_1-norm and Related Methods, Ed. Y. Dodge, North-Holland, (1987), 83–94.

[27] P. J. Huber, The 1972 wald memorial lectures, Ann. Statist., **1** (1973), 799–821.

[28] A. V. Ivanov, Consistency and asymptotic normality of least absolute value estimates, Ukranian Math. J., **36** (1984), 267–272.

[29] L. A. Jackel, Estimating regression coefficients by minimizing the dispersion of the residuals. Ann. Math. Statist., **43** (1972), 1449–1458.

[30] R. Koenker and G. Bassett, Tests of linear hypotheses and L_1 estimation, Econometrica, **50** (1982), 1577–1583.

[31] H. L. Koul, G. L. Sievers and J. W. McKean, An estimator of the scale parameter for the rank analysis of linear models under general score functions, Scand J. Statist., **14** (1987), 131–141.

[32] Z. Y. Li, An asymptotic optimality of the discriminant procedure based on a nearest neighbor density estimator, Hiroshima Math. Journal, **22** (1992), 649–655.

[33] Z. Y. Li, G. X. Chai and P. Sun, Least squares estimation in a partial linear model, Tech. Report, No. **303** (1992), Statist. Research Group, Hiroshima University, Hiroshima.

[34] Y. P. Mack and Silverman, Weak and strong uniform consistency of kernel regression estimates, Z. Wahrscheinlichkeitstheorie verw. Gebiete, **61** (1982), 405–415.

[35] T. Mathew and K. Nordstrom, Least squares and least absolute deviation procedures in approximately linear models, Statistics & Probability Letters, **16** (1993), 153–158.

[36] J. W. McKean and R. M. Schrader, Least absolute errors analysis of variance, Statistical Data Analysis Based on the L_1-norm and Related Methods, North Holland (1987), 297–305.

[37] W. Obenhofer, The consistency of nonlinear regression minimizing the L_1 norm, Ann. Statist., **10** (1982), 316–319.

[38] E. Parzen, On the estimation of a probability density function and the mode, Ann. Math. Statist., **33** (1962), 1065–1076.

[39] J. L. Powell, Least absolute deviations estimation for the censored regression model, J. Econometrics, **25** (1984), 303–325.

[40] J. L. Powell, Censored regression quantiles, J. Econometrics, **32** (1986), 143–155.

[41] B. L. S. P. Rao, Nonparametric Functional Estimation, Academic Press, London, 1983.

[42] C. R. Rao, Methodology based on the L_1 norm, in statistical inference, Sankhya, **50** (1988), 289–313.

[43] C. B. Richard, Asymptotic normality of some kernel-type estimators of probability density, Statistics & Probability Letters, **1** (1983), 295–300.

[44] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, Ann. Math. Statist., **27** (1956), 832–837.

[45] G. G. Roussas, Asymptotic normality of the kernel estimate under dependence conditions: application to Hazard rate, J. Statist. Planning and Inference, **25** (1990), 81–104.

[46] R. J. Serfling, Approximation Theorem of Mathematical Statistics, John Wiley & Sons, Inc. (1980), 95.

[47] J. B. Steven and P. T. Chris P. Tsokos, Developments in nonparametric density estimation, International Statist. Review., **48** (1980), 267–287.

[48] C. Stone, Consistent nonparametric regression, Ann. Statist., **5** (1977), 595–645.

[49] K. L. Tsui, N. P. Jewell and C. F. J. Wu, A nonparametric approach to the truncated regression problem, J. Amer. Statist. Assoc., **83** (1988), 785–792.

[50] V. N. Vapnik and A. Ya. Chervonenkis, On the uniform convergence of the relative frequencies of events to their probabilities, Theory of Prop. and Its Application, **16** (1971), 264–280.

[51] W. Y. Zhang, On the congruent kernel estimate of error distribution in linear model, J. Sichuan Univ. Natural Sci. Ed., **27** (1990), 132–144.

[52] L. C. Zhao and G. X. Chai, Survey of Chinese work on nonparametric density estimation, regression and discrimination, The Development of Statistics: recent contributions from China, Edit by X. R. Chen, X. R. *et al*, Pitman Research Notes in Mathematics Series, **258** (1992), 199–217.

[53] Y. Vardi, Empirical distribution in selection bias models, Ann. Statist., **13** (1985), 178–205.

[54] M.-C. Wang, N. P. Jewell and W.-Y. Tsai, Asymptotic properties of the product limit estimate under random truncation, Ann. Statist., **14** (1986), 1597–1605.

[55] M. Woodroofe, Estimating a distribution function with truncated data, Ann. Statsit., **13** (1985), 163–177.

[56] Y. Wu, Strong consistency and exponential rate of the 'Minimum L_1-norm' estimate in linear regression model, Computational Statist. and Data Analysis, **6** (1988), 285–295.

Department of Mathematics
Faculty of Science
Hiroshima University
and
Department of Mathematics
Sichuan University